Credit EDA Case Study

• • •

Prajwal
Tushar Matlani

Introduction

This case study aims to apply EDA in a real business scenario of risk analytics in banking and financial services and understand how to minimize the risk of losing money while lending to customers

Business Objective

The case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as

- Loan Denial
- Reducing Loan Amount
- Lending at Higher Interest

To understand the driving factors behind loan default, i.e. the variables which are strong indicators of default and utilise this knowledge for its portfolio and risk assessment.

Analysis Approach

Data Cleaning
- Handling missing and null values
- Handling Data Imbalance and Outliers

Univariate Analysis
- Analysis of Categorial Variables
- Analysis of Numerical Variables

Bivariate Analysis
- Analysis of Categorial and Numerical Variables
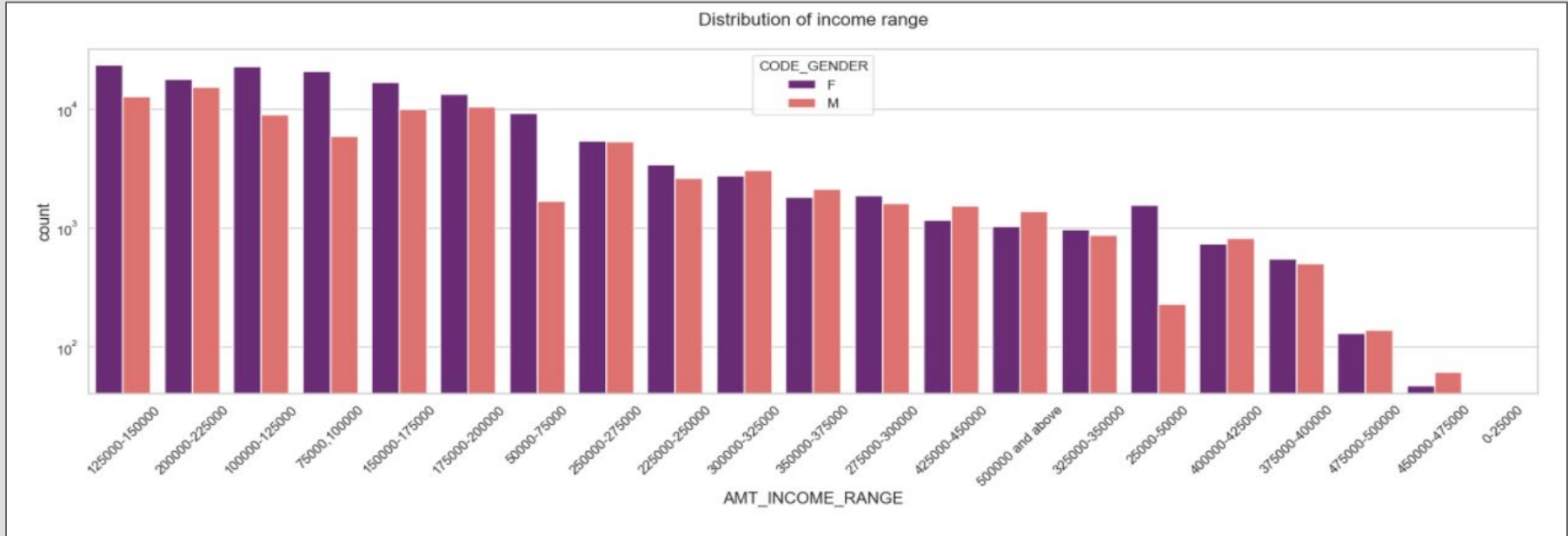- Analysis on correlated data

Insight and Conclusion
- Inferences based on correlated data sets
- Suggestions based on combined data sets

Analysis for Target 0
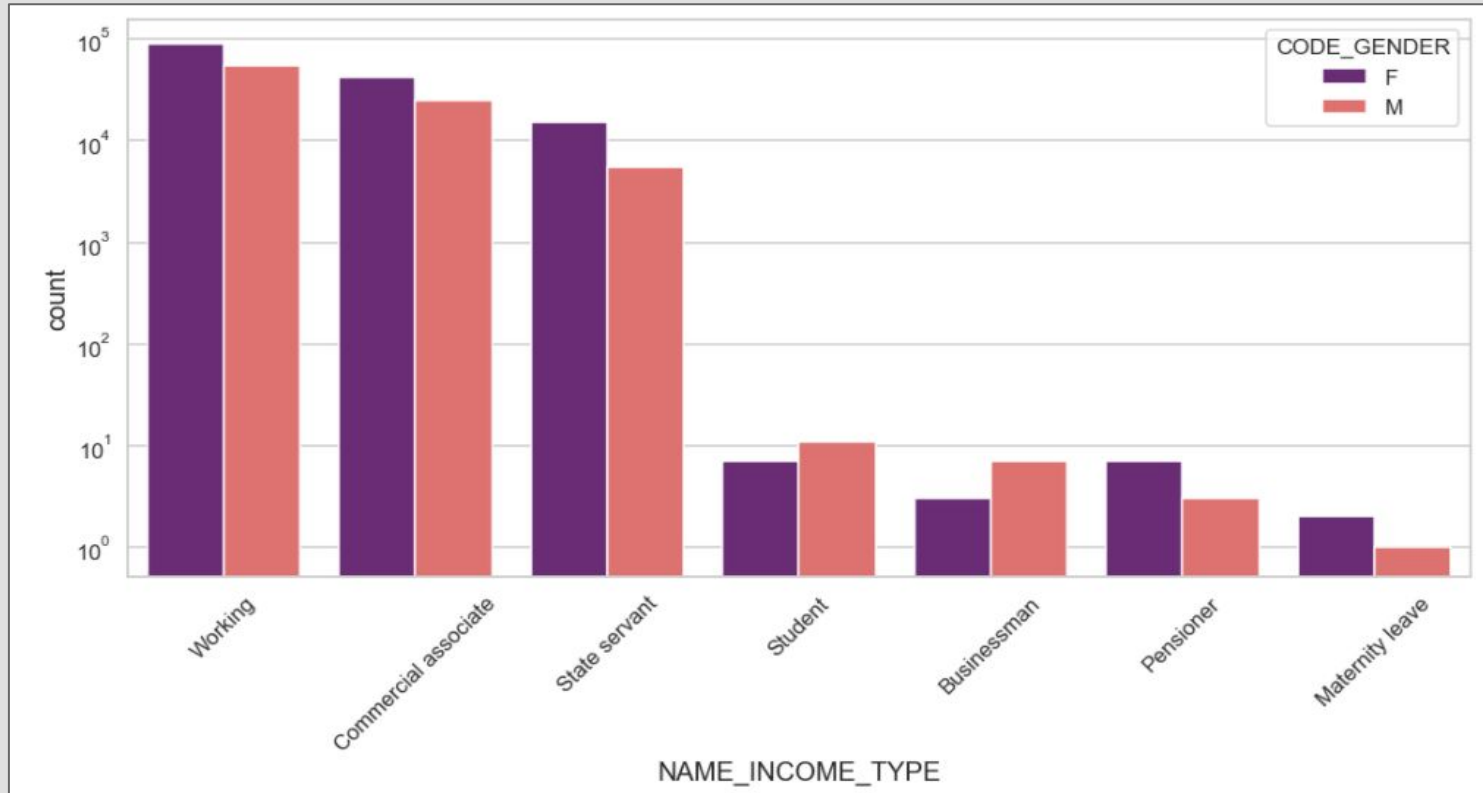## (Client with no payment difficulties)

Categorical Univariate Analysis- Target 0

Distribution of income range based on gender
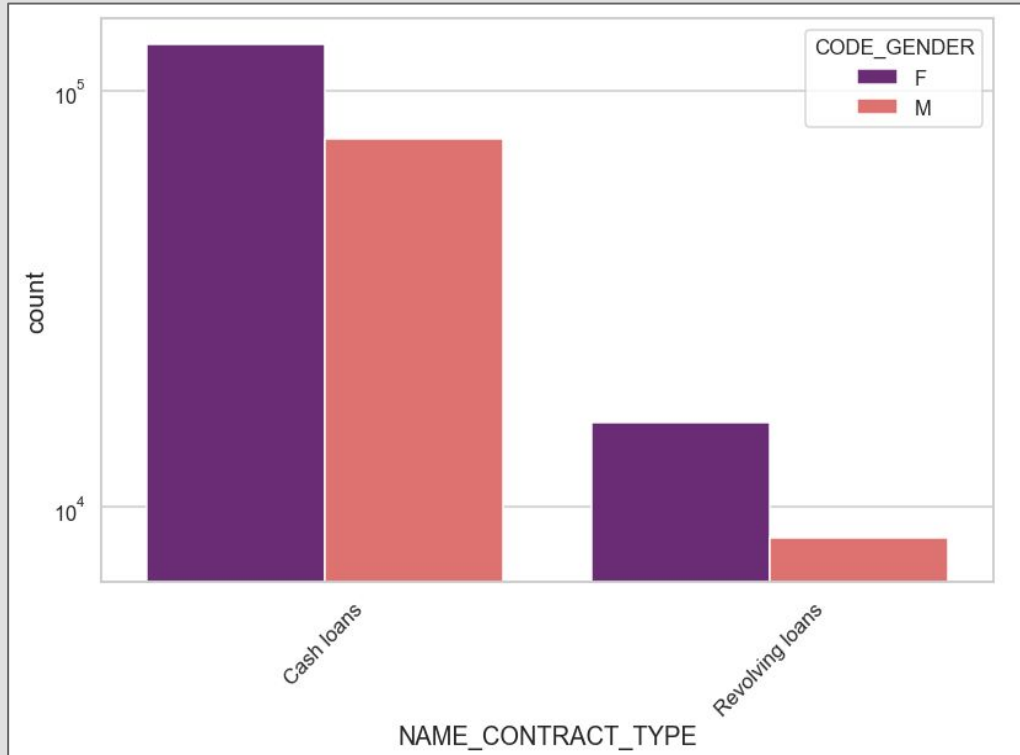

Distribution of income range

- Female counts are higher than male.
- Income range from 100000 to 200000 is having more number of credits.
- This graph show that females are more than male in having credits for that range.
- Very less count for income range 400000 and above

Distribution of Income Type

- For income type 'working', 'commercial associate', and 'State Servant' the number of credits are higher than others.
- For this Females are having more number of credits than male.
- Less number of credits for income type 'student' ,'pensioner', 'Businessman' and 'Maternity leave'.
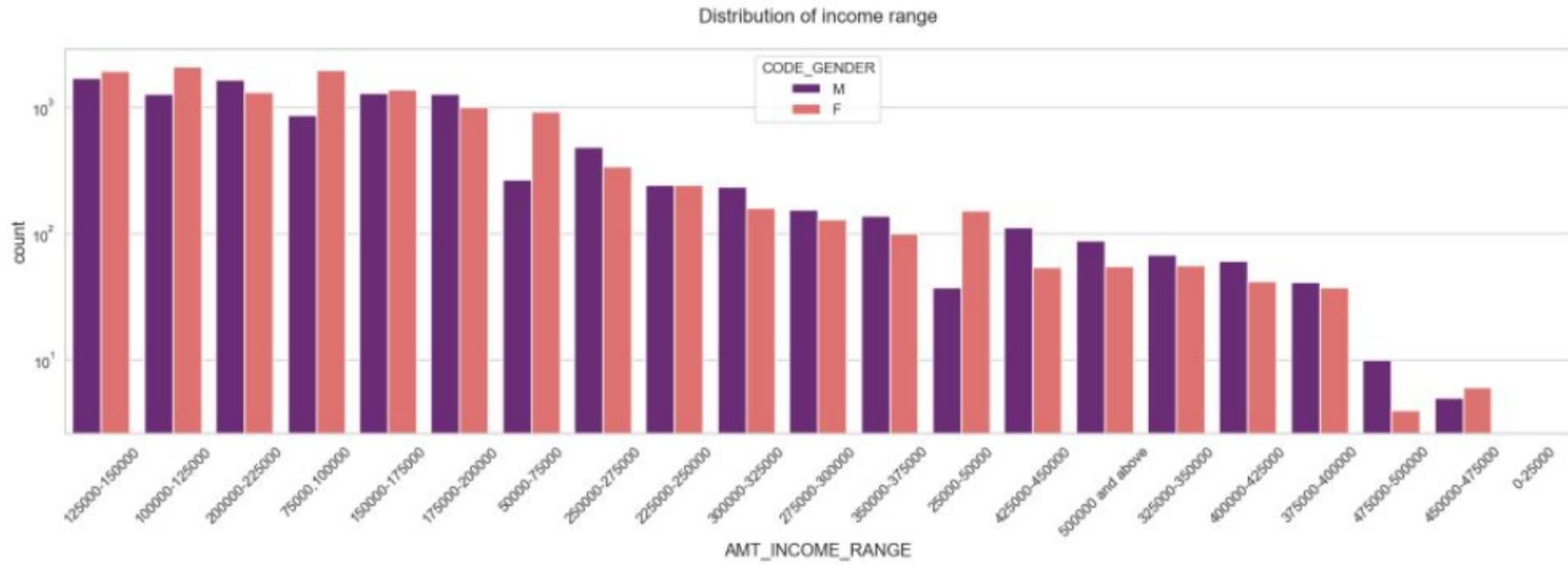
Distribution of Contract Type

- Contract type 'cash loans' has higher number of credits than 'Revolving loans'
- More Female applying credit than Male

Analysis for Target 1
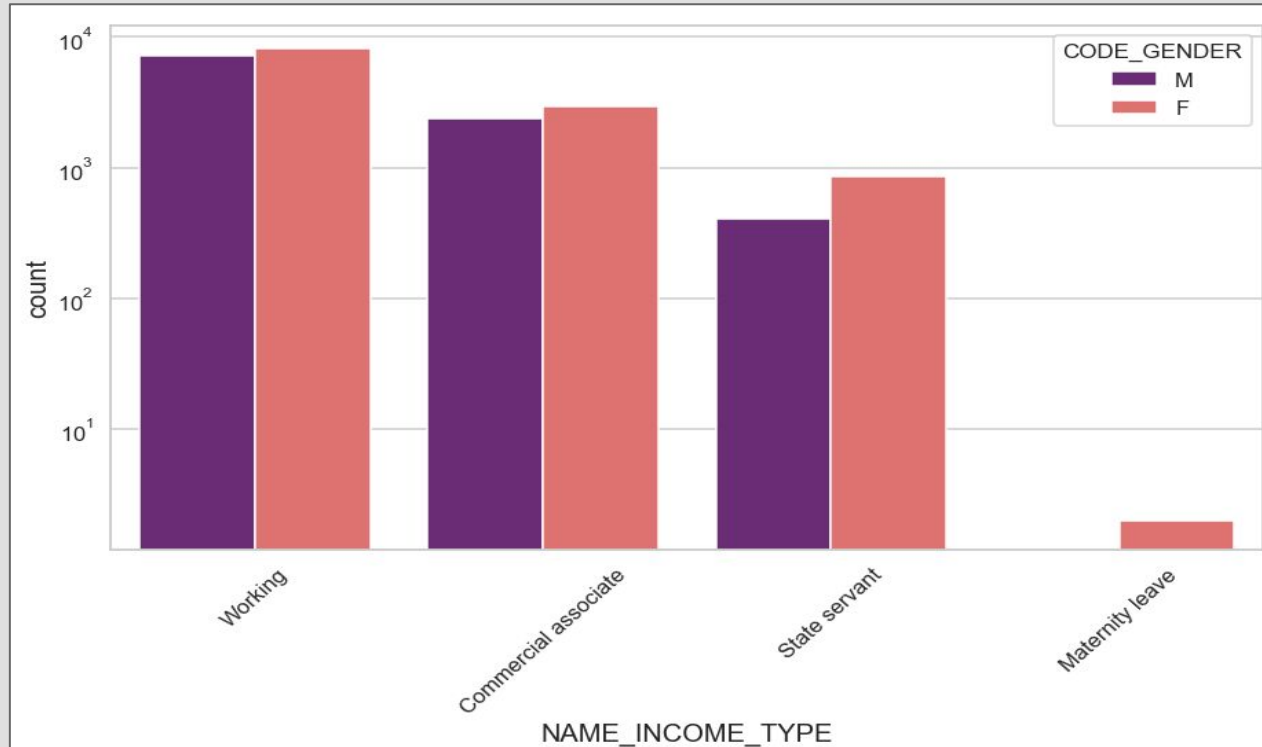# (Client with payment difficulties)

Categorical Univariate Analysis- Target 1

Distribution of Income Range

Distribution of income range

- Count for Male is more than female in having credits for range 100000 to 200000
- Low count for income range 400000 and above

Distribution of Income Type

- For income type 'working', 'commercial associate', and 'State Servant' the number of credits are higher than other i.e. 'Maternity leave.
- Females have more number of credits than male.
- Less number of credits for income type 'Maternity leave'
- For type 1: There is no income type for 'student', 'pensioner' and 'Businessman' which means they don't do any late payments.
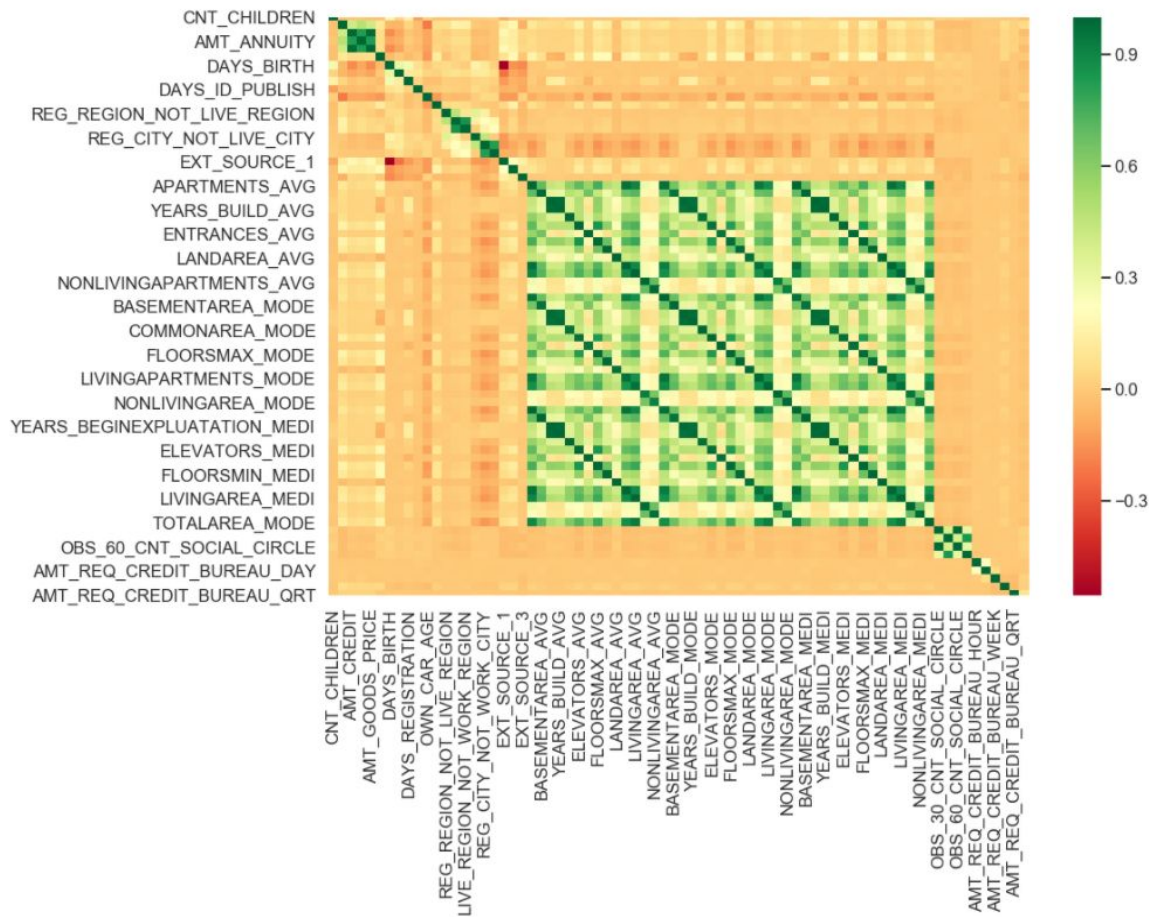
Distribution of Contract Type
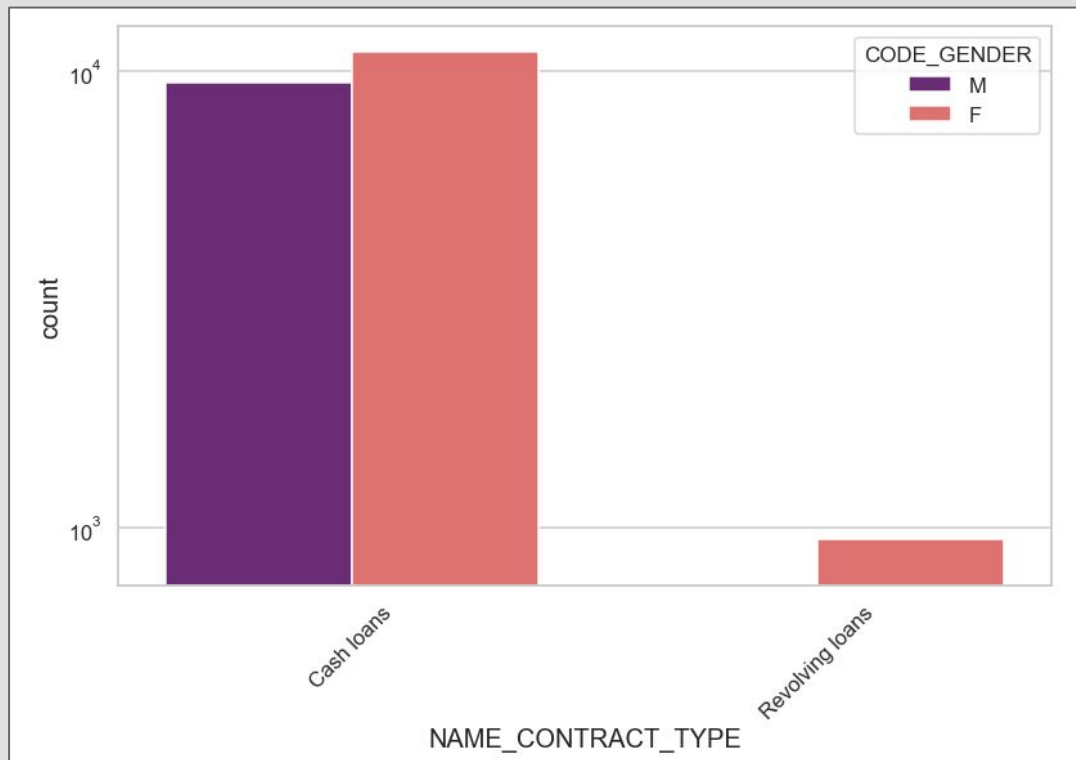
- For contract type 'cash loans' is having higher number of credits than 'Revolving loans' contract type

- For this also Female is leading for applying credits

- For type 1 : there is only Female Revolving loans

Correlation of Numerical Columns for both Target 0 and 1 represented through heat map

Correlation for Target 0

- Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
- Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.
- Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.
- less children client have in densely populated area.
- Credit amount is higher to densely populated area.
- The income is also higher in densely populated area

Correlation for Target 1
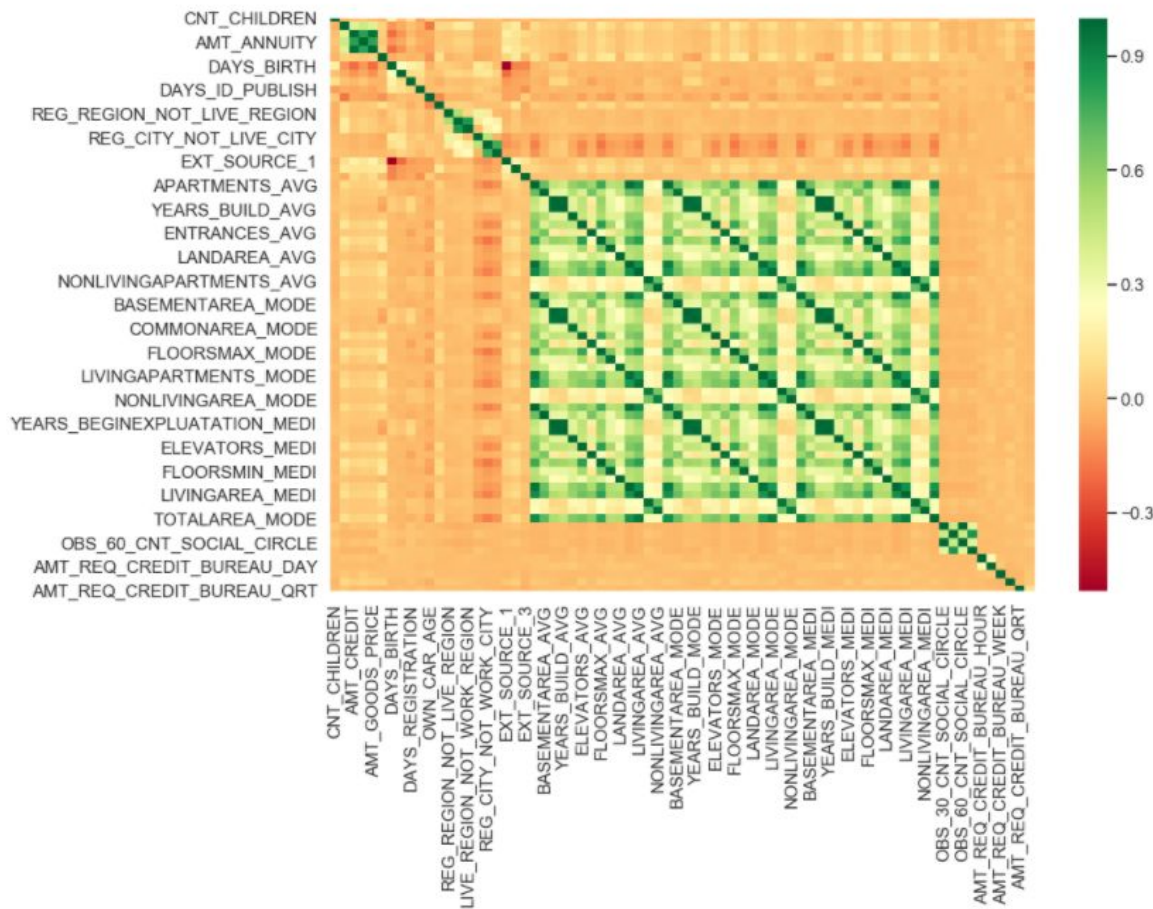
This heat map for Target 1 has a similar observation like Target 0. But there are some differences mentioned below

- The client's permanent address does not match contact address are having less children and vice-versa
- The client's permanent address does not match work address are having less children and vice-versa

Numerical Bivariate Analysis


Credit amount vs Education Status
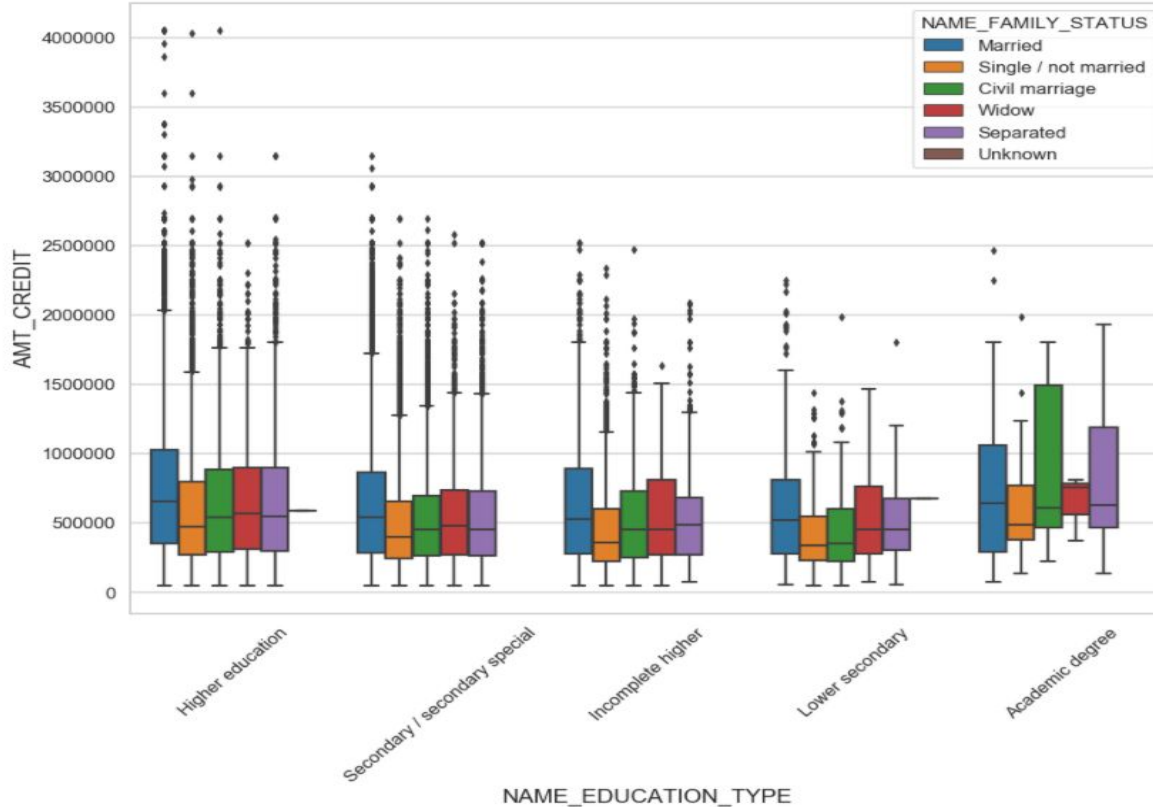
Box Plotting for Credit Amount and Education Status

From the above box plot we can conclude that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.

Also, higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers. Civil marriage for Academic degree is having most of the credits in the third quartile.

Numerical Bivariate Analysis
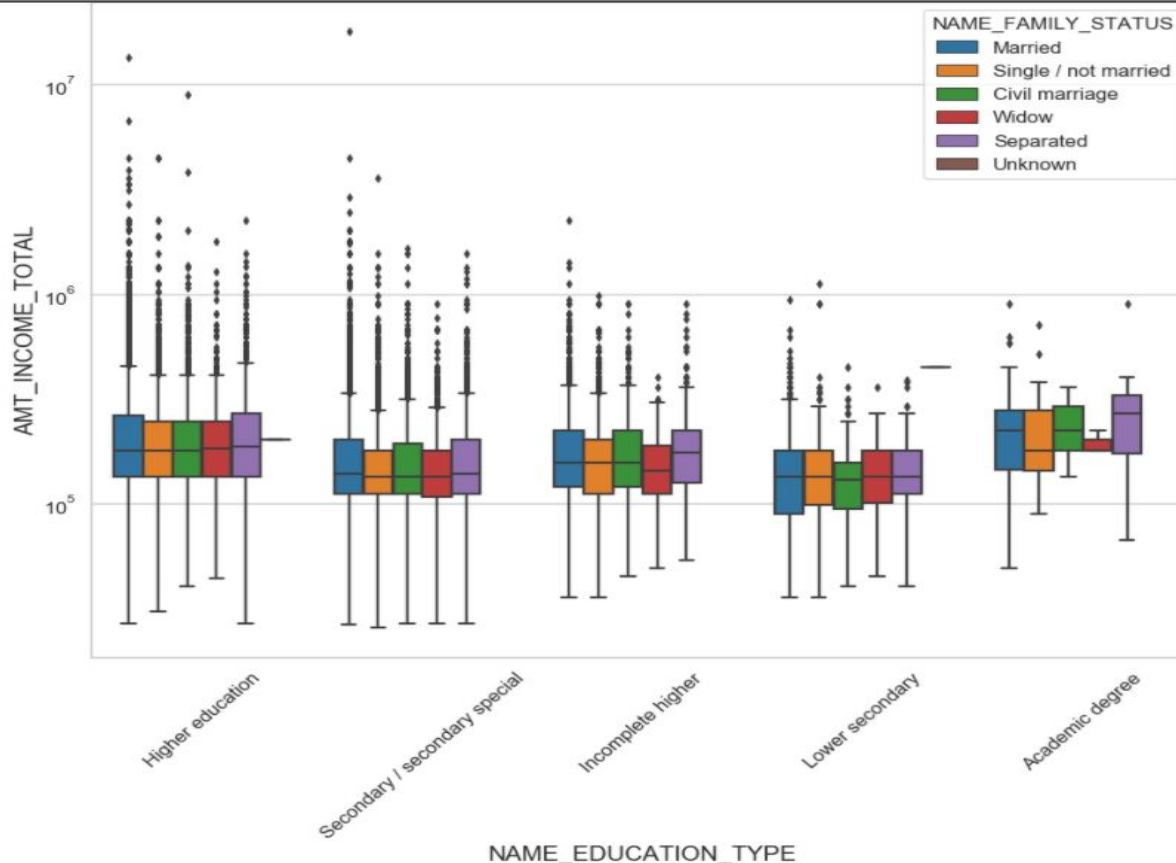
Box Plotting for Income and Education Status

From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.
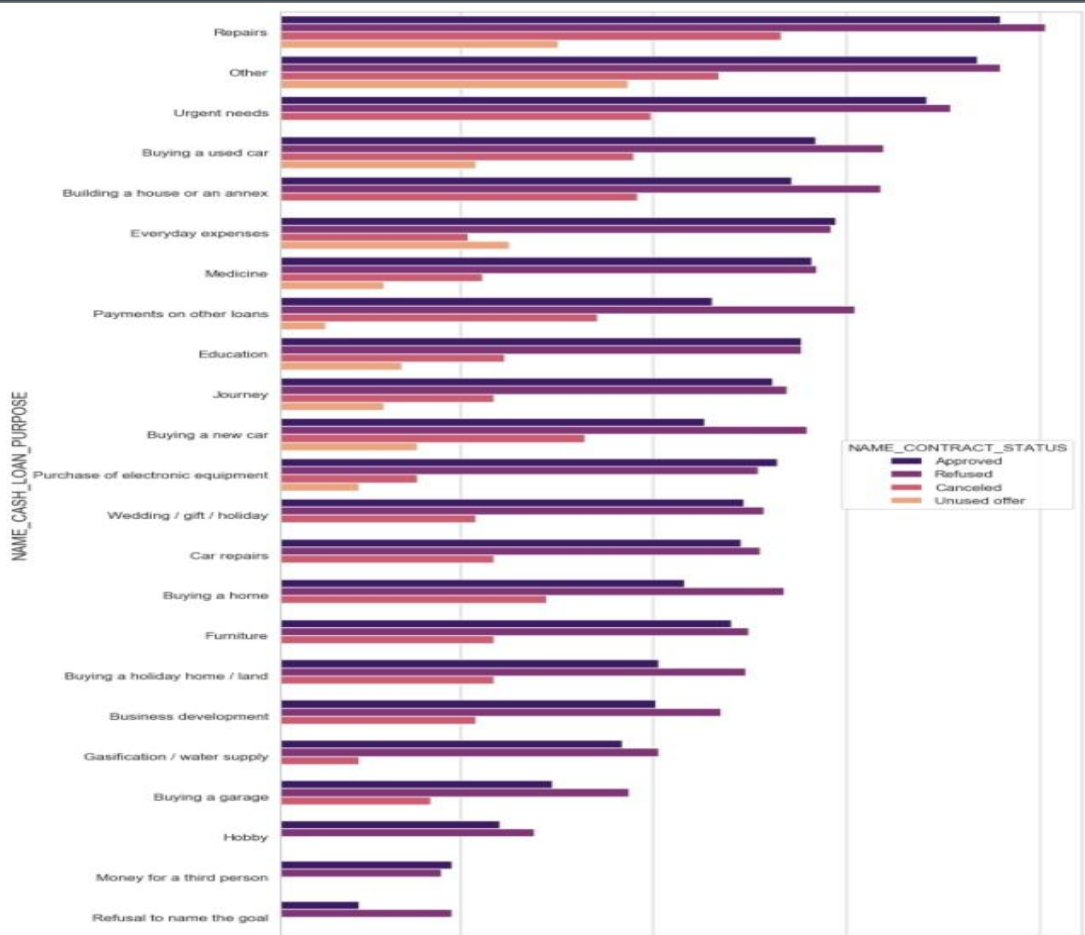
It does contain many outliers. Less outlier are having for Academic degree but their income amount is little higher that Higher education.

Lower secondary of civil marriage family status are have less income amount than others.

Analysis on Combined Data
Univariate and Bivariate
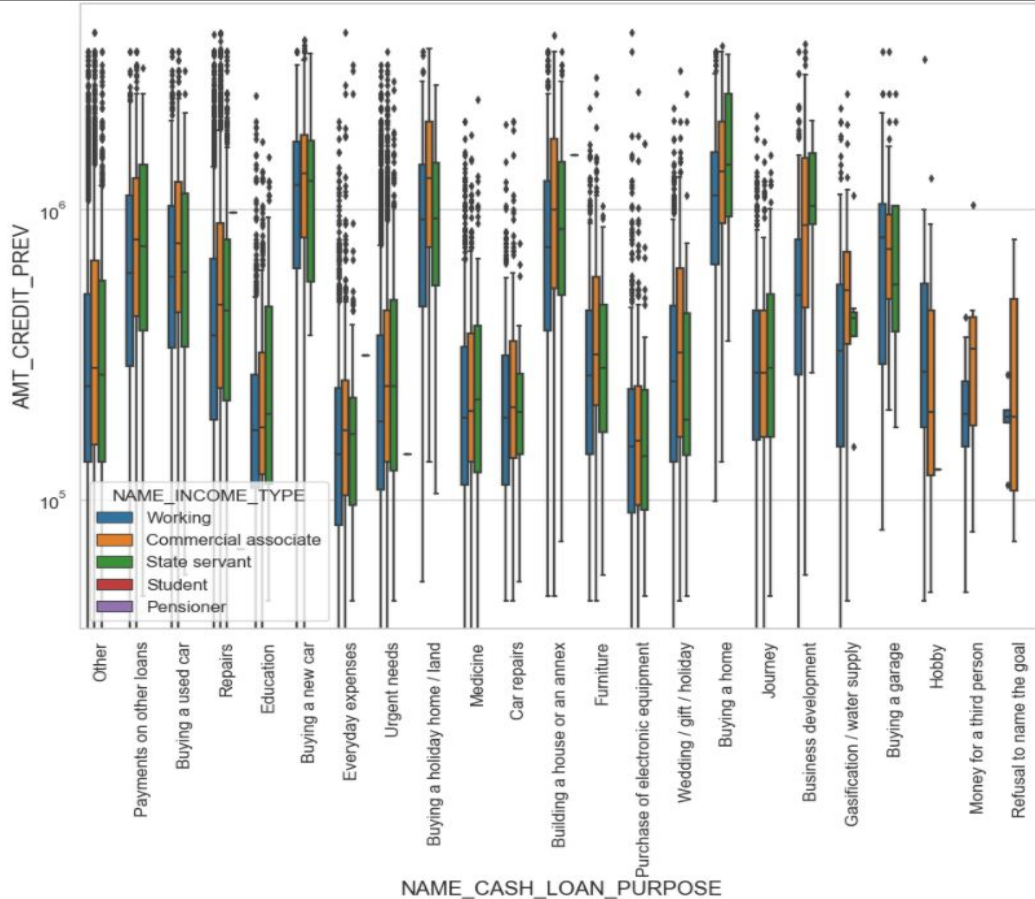
Univariate Analysis on Combined Data

Distribution of Contract Status with Purpose

Most rejection of loans came from purpose 'repairs'.

For education purposes we have equal number of approves and rejection

Paying other loans and buying a new car is having significant higher rejection than approves

Bivariate Analysis on Combined Data

Boxplotting for Credit Amount and Loan Purpose

- The credit amount of Loan purposes like 'Buying a home','Buying a land', 'Buying a new car' and 'Building a house' is higher.

- Income type of state servants have a significant amount of credit applied

- Money for third person or a Hobby is having less credits applied for.

Bivariate Analysis on Combined Data
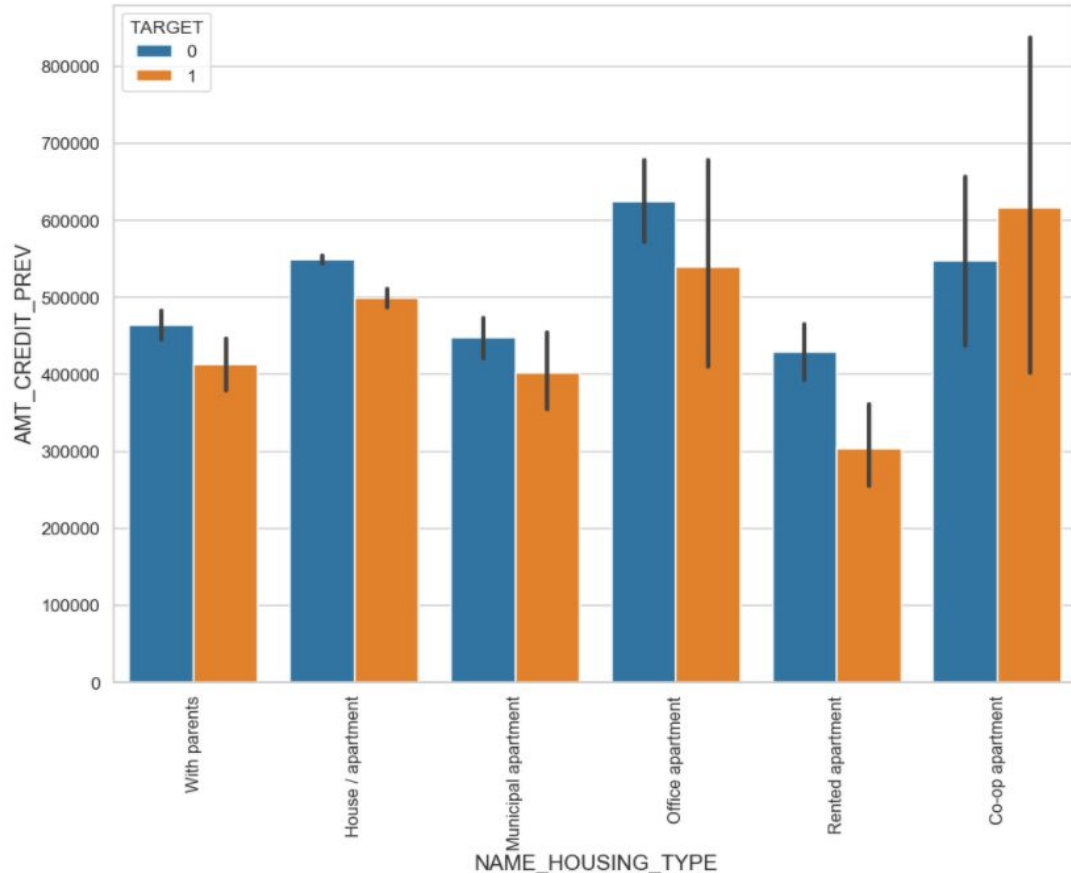
Boxplotting for Previous Credit Amount and Housing Type

Here for Housing type, office apartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1.

So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment. Bank can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments.

Conclusive Suggestions

- Banks should focus more on contract type 'Student', 'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.

- Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.

- Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.

- Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments